CENTRE DE RECHERCHE SUR LES TRANSPORTS
(C.R.T.)
CENTRE FOR RESEARCH ON TRANSPORTATION

UNIVERSITÉ DE MONTRÉAL
ÉCOLE DES HAUTES ÉTUDES COMMERCIALES
ÉCOLE POLYTECHNIQUE

CENTRE DE RECHERCHE SUR LES TRANSPORTS
UNIVERSITÉ DE MONTRÉAL
C.P. 6128, SUCCURSALE A.
MONTRÉAL
CANADA H3C 3J7

TÉLÉPHONE: (514) 343-7575
TÉLEX: (21) 05-561325
EMAIL: crt@crt.umontreal.ca
TÉLÉCOPIEUR (FAX): (514) 343-7121

91-06708/

CRT-794

The Nonlinear Bilevel Programming Problem: A General Formulation and Optimality Conditions

by

Yang Chen[1,2]
Michael Florian[1,2]

[1] Centre de recherche sur les transports, Université de Montréal, C.P. 6128, succursale A, Montréal, Canada H3C 3J7.

[2] Département d'informatique et recherche opérationnelle, Université de Montréal, C.P. 6128, succursale A, Montréal, Canada H3C 3J7.

Centre de recherche sur les transports — Publication #794
Département d'informatique et de recherche opérationnelle — Publication #782

Juillet 1991

Dépôt légal – 3ème trimestre 1991

ABSTRACT

The bilevel programming problem (BLPP) is characterized by a sequence of two optimization problems where the constraint region of the first is determined implicitly by the solution to the second. It is well known that the problem is "hard" due to its inherent nonconvexity and nondifferentiability. The algorithms that have been stated for the BLPP are limited to special cases of the problem. Also, to date, necessary optimality conditions have not been stated for the general BLPP. In this paper, we present a general mathematical programming formulation for the BLPP which is based on the marginal (value) function. This function is defined to be the infimum value function of the second level problem, parametrized by the first level variables. We investigate the properties of the new formulation in the framework of nondifferentiable optimization and analyze the (generalized) differentiabilities of the marginal function under fairly general conditions. The first order necessary optimality conditions for the BLPP are developed and, consequently, the cone of the feasible direction of the BLPP is characterized. Both are important as the basis for designing efficient algorithms for solving the general BLPP.

Keywords: bilevel programming; marginal function; set-valued map; optimality and stability; nondifferential optimization.

* * * * * * * * *

RÉSUMÉ

Le problème de programmation mathématique bi-niveaux (PPBN) est caractérisé par une suite de deux problèmes d'optimisation dont le domaine réalisable du premier problème est déterminé implicitement par la solution du deuxième. C'est bien connu que ce problème est "difficile" à cause du fait qu'il n'est pas convexe ni différentiable. Les algorithmes dévelopés pour le PPBN sont limités à résoudre de cas particuliers de ce problème. À date, les conditions nécessaires d'optimalité n'ont pas encore été énoncées pour le PPBN général. Dans cet article nous présentons une formulation de programmation mathématique pour le PPBN général basé sur la fonction marginale. Cette fonction est définie comme étant l'infinimum du problème du deuxième niveau paramétrisé par les variables du premier niveau. Nous analysons les propriétés de cette nouvelle formulation dans le cadre de l'optimisation non différentiable et, en particulier, la différentiabilité de la fonction marginale. Les conditions nécessaires d'optimalité du premier ordre pour le PPBN sont développées et, par conséquent, le cône des directions réalisables est caractérisé. Ces résultats peuvent servir pour le développement des algorithmes efficaces pour la résolution du problème PPBN général.

Mots-clés: programmation mathématique bi-niveau; applications multivoques; optimalité et stabilité; optimisation non différentiable.

1 Introduction

The bilevel programming problem (BLPP), which is characterized by a set of two optimization problems where the constraint region of the first is determined implicitly by the solution to the second, has been dedicated much attention by current research during the last decade. The problem may be stated mathematically as follows:

$$
\begin{aligned}
(BLPP) \qquad & \min F(x,y) \\
& s.t.\ \ l(x,y) \leq 0, \\
& \qquad y = \arg\min f(x,y) \\
& \qquad\qquad s.t.\ \ g(x,y) \leq 0,
\end{aligned}
$$

where the variables x and y are controlled by the first and second levels respectively. A large variety of real world problems, where hierarchical decision-making processes are involved, can be formulated as multilevel or bilevel problems. A Stackelberg duopoly, or leader- follower continuous game, studied by Von Stackelberg (1952) for static economic problems, and for dynamic version by Chen and Cruz (1972), is the classical example of BLPP. Other examples include government regulation (Candler and Townsley (1982), for instance), central economic planning and management of a decentralized firm. Several bilevel programming problems originate in the transportation (Marcotte (1986), LeBlanc and Boyce (1986), Friesz et al. (1990) and Florian and Chen (1991)). The max-min problem (Danskin (1966), Falk (1972)) is also a special case of the BLPP.

The linear BLPP has been studied extensively by many authors. Bialas and Karwan (1982, 1984) and Candler and Townsley (1982) investigate the structure of the linear BLPP thoroughly and develope some algorithms based on the idea of exploring among the extreme points of the induced region of the problem. Fortuny-Amat and McCarl (1981), Bard and Falk (1982), Bard and Moore (1990) and Hansen, Jaumard and Savard (1990) develop branch- and-bound algorithms to solve the BLPP, in which branching is undertaken on the complementarity conditions derived from applying the Karush-Kuhn-Tucker condition to the second level problem. Judice and Faustino (1988, 1989) propose the use of parametric complementary pivoting algorithms, which pro-

vide an ϵ-optimal solution for any given ϵ . Benson (1989) studies the structure and properties of the linear multilevel programming problem (LMPP) that may be unbounded and points out the relation between the (LMPP) and a certain parametric concave minimization problem.

For the nonlinear BLPP, Bard and Falk (1982) and Bard (1983) study some properties of the BLPP and a one- dimensional search algorithm that yields a locally optimal solution. Aiyoshi and Shimisu (1981) transform the BLPP into a one level optimization problem, whose constraints consist of the necessary optimality conditions of the penalized second level problem under several strong assumptions, and solve it by a penalty function method. Kolstad and Lasdon (1990) adapt the methods and theory from sensitivity analysis, presented in Fiacco (1976), to evaluate the derivative with respect to the y variables as an implicit function of the x-variables and use a descent approach to handle the problem. Gauvin and Savard (1990) propose a steepest descent direction method to solve the nonlinear BLPP, where the descent direction of the first level is determined by the optimal solution of another quadratic BLPP, which is easier to solve than the original problem. Loridan and Morgan (1989a, 1989b) suggest an approximation scheme to deal with the BLPP and get some interesting results, particularly, an existence result for the solution to the ϵ- regularized problem, which is a relaxation of some of the problem conditions.

In general, the BLPP is intractable because of its inherent nonconvexity and nondifferentiability. Even the simplest case, the linear BLPP, has been shown to be NP-hard (Jeroslow, 1985). Therefore, it seems that to find efficient global algorithms for the general BLPP is a formidable task; partly because of the nonconvexity of the problem and, in the authors' opinion, partly because of the lack of sophisticated theoretical analysis of properties, when compared with results available for standard (one level) mathematical programming problems. Bard (1984) proposes a set of conditions as the necessary optimality conditions for the BLPP. Unfortunately, a counterexample for the conditions is provided by Clark and Westerberg (1988). To the best of the authors' knowledge, necessary optimality conditions for the general BLPP are not available to date. Another feature of the existing solution proce-

dures for the BLPP is that most of these are based on the formulation which is obtained by applying the KKT condition to the second level problem. It is important to point out that, from the theoretical point of view, this KKT formulation of the nonlinear BLPP is not equivalent in general to the BLPP. Also, from a computational point of view, the KKT conditions may involve the addition of a large number of variables and complicated constraints to the first level problem, if the second level problem is of large scale, such as bilevel problems where the second level problem is a network equilibrium problem (Marcotte (1986), Ben-Ayed et al. (1988), Spiess (1990), Friesz et al. (1990) and Florian and Chen (1991)). Thus it is important to find alternative equivalent formulations of the BLPP in order to develop the necessary theory in order to develop efficient algorithms for it.

The purpose of this paper is to establish the first order optimality conditions for the BLPP. To this end, we present first a general mathematical programming formulation of the BLPP, which is based on the marginal (value) function. This function is defined to be the infimum value function of the second level problem, parametrized by the first level variables. We investigate the properties of this new formulation in the framework of nondifferentiable optimization and analyse the (generalized) differentiabilities of the marginal function under fairly general conditions. The first order necessary optimality conditions for the BLPP are developed and, consequently, the cone of the feasible direction of the BLPP is characterized. Both are important for designing efficient algorithms for solving the general BLPP.

The paper is organized as follows. After this introduction, in section 2, some basic properties of the marginal function are reviewed, the general formulation of the BLPP based on the marginal function is presented and its properties are discussed. Section 3 is devoted to the analysis of the (generalized) differentiability of the marginal function under fairly general conditions. The first order necessary conditions and the feasible direction cone of the BLPP are established in section 4 and are applied to a "special" example. In section 5, more simpler and powerful necessary conditions are derived for the convex BLPP by applying the general results to it and the generalized necessary

optimality conditions for the min-max problem are obtained as well. Finally, some future research directions on the BLPP conclude the paper briefly in section 6.

2 Preliminaries and Problem Formulation

The nonlinear bilevel programming problem (BLPP) is defined as follows:

$$(P1) \qquad \min_{(x,y)\in X\times Y} F(x,y) \tag{1}$$

$$s.t. \quad l(x,y) \leq 0, \tag{2}$$

$$y \in \arg\min_{y\in Y} f(x,y) \tag{3}$$

$$s.t. \quad g(x,y) \leq 0, \tag{4}$$

where X and Y are two compact subspaces, $l(\cdot,\cdot): X \times Y \to \mathbf{R}^s$ and $g(\cdot,\cdot)$: $X \times Y \to \mathbf{R}^t$ are vector functions and $F(\cdot,\cdot)$: $X \times Y \to \mathbf{R}$ and $f(\cdot,\cdot): X \times Y \to \mathbf{R}$ are objective functions of the first and second level respectively. Level one objective function $F(\cdot,\cdot)$ and constraints $l(\cdot,\cdot)$ are assumed to be once continuously differentiable and the second level objective function $f(\cdot,\cdot)$ and the constraints $g(\cdot,\cdot)$ are assumed to be twice continuous differentiable. In addition, $f(\cdot,\cdot)$ is assumed to be locally Lipschitz.

The following notation will be used in the sequel.

Constraint region of BLPP:

$$S = \{(x,y) \mid x \in X,\ y \in Y,\ l(x,y) \leq 0,\ g(x,y) \leq 0\}.$$

Feasible region of the second level problem for fixed $x \in X$:

$$Y(x) = \{y \in Y \mid g(x,y) \leq 0\}.$$

Rational reaction set of the second level problem:

$$O(x) = \{y \in Y \mid y \in \arg\min_{z\in Y(x)} f(x,z)\}$$

Feasible region of BLPP:

$$\overline{S} = \{(x,y) \mid (x,y) \in S,\ y \in O(x)\}$$

The feasible region $Y(x)$ and the rational reaction set $O(x)$ define two set-valued maps from X to Y.

A solution (x^*, y^*) is said to be optimal for the BLPP if,
(i) it is feasible, i.e. $(x^*, y^*) \in \overline{S}$ and,
(ii) $F(x^*, y^*) \leq F(x, y)$ for all feasible (x, y).

Clearly the second level problem can be viewed as a parametric mathematical programming problem parametrized by the first level variable x:

$$(P_x) \qquad \min f(\bar{x}, y)$$

$$s.t.\ \ g(\bar{x}, y) \leq 0.$$

that gives rise naturally to a marginal (value) function associated with its infimum value. It is this fact and results from parametric mathematical programming that motivate us to have a new look at the structure and properties of the BLPP. The marginal function $v(\cdot) : X \to \mathbf{R}$ is then defined as

$$v(x) = \inf_{y \in Y(x)} f(x,y). \tag{5}$$

The rational reaction set $O(x)$ may be rewritten as

$$O(x) = \{y \in Y \mid f(x,y) = v(x)\}$$

and is also called the *marginal map*. The marginal function $v(x)$ and the marginal map $O(x)$ will be shown later to play a crucial role in the new formulation of the BLPP. Thus it is important to study extensively their analytical properties. Several authors have contributed many results on the properties of set-valued maps and marginal functions (See Hogan (1973), Gauvin and Dubeau (1982) and Rockafellar (1984)). The following propositions focus on the (semi) continuity of the marginal function $v(x)$, which can be considered as corollaries of results given by Hogan (1973).

Proposition 1 *If $f(\cdot,\cdot)$ is lower semicontinuous (l.s.c.) on $X \times Y$, each component of $g(\cdot,\cdot)$ is l.s.c. on $x_0 \times Y$ and if $Y(x_0)$ is compact, then $v(\cdot)$ is l.s.c. at x_0.*

Proof. Since f is l.s.c. on $X \times Y$, then for any $y' \in Y$ one can associate neighborhoods $\mathcal{N}_{y'}(x_0)$ of x_0 and $\mathcal{N}(y')$ such that for any $x \in \mathcal{N}_{y'}(x_0)$, $y \in \mathcal{N}(y')$

$$f(x,y) \geq f(x_0,y') - \epsilon/2 \qquad for\ \forall \epsilon > 0. \tag{6}$$

$Y(x_0)$ can be covered by n neighborhoods $\mathcal{N}(y_i)$ because of the compactness of $Y(x_0)$. Then

$$\mathcal{N} = \bigcup_{i=1}^{n} \mathcal{N}(y_i)$$

is a neighborhood of $Y(x_0)$. Since each component of $g(\cdot,\cdot)$ is l.s.c., thus $Y(\cdot)$ is upper semicontinuous (u.s.c.) (Theorem 10, Hogan (1973)). Therefore there exists a neighborhood $\mathcal{N}_0(x_0)$ of x_0 such that

$$Y(x) \subset \mathcal{N} \qquad for\ \forall x \in \mathcal{N}_0(x_0)$$

Let

$$\mathcal{N}(x_0) = \mathcal{N}_0(x_0) \bigcap \bigcap_{i=1}^{n} \mathcal{N}_{y_i}(x_0).$$

When $x \in \mathcal{N}(x_0)$ and $y \in Y(x)$, then y belongs to $\mathcal{N}$ and thus to some $\mathcal{N}(y_i)$. Hence, by (6),

$$f(x,y) \geq f(x_0,y_i) - \epsilon/2 \geq v(x_0) - \epsilon.$$

Therefore

$$v(x) \geq v(x_0) - \epsilon. \tag{7}$$

If $Y(x) = \emptyset$ for $\forall x \in \mathcal{N}(x_0)$, then $v(x) = +\infty$. (7) still hold. Therefore $v(\cdot)$ is l.s.c. at x_0.. □

Proposition 2 *If $f(\cdot,\cdot)$ is u.s.c. on $X \times Y$ and $Y(\cdot)$ is l.s.c. at x_0, then $v(\cdot)$ is u.s.c. at x_0.*

Proof. For $\epsilon > 0$, there exists $y_0 \in Y(x_0)$ such that

$$v(x_0) \geq f(x_0, y_0) - \epsilon/2.$$

Since f is u.s.c., then $\exists$ neighborhoods $\mathcal{N}_1(x_0)$ of x_0 and $\mathcal{N}(y_0)$ of y_0 such that

$$f(x, y) \leq f(x_0, y_0) + \epsilon/2$$

for $\forall x \in \mathcal{N}_1(x_0)$ and $\forall y \in \mathcal{N}(y_0)$. Because of the lower semi continuity of $Y(.)$ at x_0, there exists a neighborhood $\mathcal{N}_2(x_0)$ such that

$$Y(x) \cap \mathcal{N}(y_0) \neq \emptyset \quad for\ \forall\ x \in \mathcal{N}_2(x_0).$$

So for $\forall\ x \in \mathcal{N}_1(x_0) \cap \mathcal{N}_2(x_0)$, there exists $y \in Y(x) \cap \mathcal{N}(y_0)$ such that

$$f(x, y) \leq f(x_0, y_0) + \epsilon/2 \leq v(x_0) + \epsilon.$$

Thus

$$v(x) = \inf_{y \in Y(x)} f(x, y) \leq v(x_0) + \epsilon.$$

This proves that $v(\cdot)$ is u.s.c. at x_0. □

Proposition 3 *(Theorem 7, 8, Hogan (1973)) If $f(\cdot, \cdot)$ is continuous on $X \times Y$ and $Y(\cdot)$ is continuous with compact values, then $v(\cdot)$ is continuous and the marginal map $O(\cdot)$ is u.s.c..* □

In order to focus on the main aspect of the analysis, the following assumption is made before discussing the formulation of the BLPP and its properties:

The set-valued maps Y(x) and O(x) are *strict*, i.e. $Y(x) \neq \emptyset$, $O(x) \neq \emptyset$ for $\forall x \in X$ (Aubin and Ekeland, 1984).

The constraint $l(x, y) \leq 0$ of the first level is also omitted without loss of generality in order to simplify the presentation.

Let us consider now the following single level mathematical programming problem:

$$(P2) \qquad \min_{(x,y) \in X \times Y} F(x, y)$$

$$s.t. \quad f(x,y) - v(x) = 0, \tag{8}$$

$$g(x,y) \leq 0. \tag{9}$$

The relation between the BLPP $(P1)$ and the one level problem $(P2)$ is easy to see by the following proposition and is stated formally in:

Proposition 4 *The problem* $(P2)$ *is equivalent to the BLPP* $(P1)$.

□

Although the transformation of the BLPP to a single level problem is trivial, it is significant to help one understand and solve the BLPP, if one uses the properties of the marginal function $v(x)$ to establish the theory on which solution algorithms for the BLPP can be based. As we mentioned briefly in the previous section, the formulation $(P2)$ has the following features:

- It may be considered as a unified framework for the BLPP in the framework of nonsmooth optimization. In fact, it is the most general form for the BLPP, in the sense of being equivalent to the bilevel problem as a single level optimization problem. One need not assume the convexity of the second level problem, such as required in the KKT formulations. Theoretically, the formulation can include all forms for the second level problem, provided that it has a global optimal solution for fixed first level variables.

- Geometrically, the feasible region of the BLPP is a surface determined by $f(x,y) - v(x) = 0$ within the set $\{(x,y) \mid g(x,y) \leq 0\}$. Thus the properties of the BLPP depend on the properties of the implicit marginal function $v(x)$. It is also easy to see that the BLPP is nondifferential and nonconvex in nature, since $v(x)$ is so in general.

- This formulation possesses two advantages from the algorithmic point of view. First, it avoids applying the KKT condition to the second level problem of the BLPP to obtain an equivalent single level optimization problem. As a consequence one need not handle a large number of extra variables and more complicated constraints introduced by the KKT formulation. Also, the

problem almost preserves the structure of the second level problem, which is very important for developing efficient algorithms in many cases. Second, it provides naturally a suitable form for decomposable solution techniques to solve the BLPP.

- In all existing approaches to solve the BLPP, multiple optima for the second level problem are excluded from consideration. However, in this formulation, one need not worry about the multiple optima of the second level problem, since there is only one objective function and it "forces" the second level problem to provide the "best" response with respect to the "single" objective function of the "entire" problem.

Clearly, the central issue in this approach is how one can use the properties of the marginal function to establish a new theoretical foundation for the BLPP and to develop efficient solution methods based on this formulation. In the following sections we will study the properties of $v(x)$ and derive necessary optimality conditions for the BLPP.

3 Differentiability of the Marginal Function $v(x)$

This section is devoted to the (general) differentiability of the marginal function. The directional derivatives and the generalized gradient of the marginal function $v(x)$ have been investigated, under different assumptions, in the literature (See for example, Gauvin and Dubeau (1982) and Rockafellar (1984)). The main result of this section may be viewed as an extension of those "classical" results in the sense that it holds under more general assumptions.

First we review some basic notions in nonsmooth analysis, which are useful in the sequel. Let a function $f : \mathbf{X} \to \mathbf{R}$ be Lipschitz near x_0. The *Clarke directional derivative* of f at x_0 in the direction $k \in X$ is defined by

$$f^0(x_0; k) = \limsup_{\substack{x \to x_0 \\ t \downarrow 0}} \frac{f(x + tk) - f(x)}{t}. \tag{10}$$

The *generalized gradient* of f at x_0, denoted $\partial f(x_0)$, is the subset of the dual space X^* given by

$$\{\xi \in X^* \mid f^0(x_0; k) \geq \xi^T \cdot k \ \textit{for all}\ k \in X\} \tag{11}$$

An extensive discussion of the properties of the generalized gradient as well as all other results in nonsmooth analysis can be found in (Clarke 1983).

Temporarily we focus our attention on the second level problem P_x. Define the first-order multiplier set of P_x associated with a feasible solution y as

$$\begin{aligned} \overset{1}{\prod}(x,y) \ &= \ \{\pi \in \mathbf{R}^m \mid \nabla_y l(x,y;\pi) = 0, \pi_i \geq 0, \\ &\qquad i = 1, \cdots, m; \ \pi_i = 0 \ if \ i \notin I(x,y)\} \end{aligned} \tag{12}$$

where $l(x,y;\pi) = f(x,y) + \pi^T g(x,y)$ is the Lagrangean of P_x, $I(x,y)$ is the index set of active constrains at y and m is the number of the components of g. To define the second-order multiplier set we adopt the notations from Rockafellar (1984). Let $\mathcal{M}(x,y)$ denote the collection of all subspaces M expressible as limits of subspaces M^j of the following form:

$$M^j \ = \ \{w \in \mathbf{R}^{|Y|} \mid w^T \cdot \nabla_y g_i(x^j, y^j) = 0, \ for\ \forall i \in I(x,y)\}, \tag{13}$$

where $(x^j, y^j) \to (x,y)$.

Then the second-order multipler set $\prod^2(x,y)$ is defined by

$$\begin{aligned} \overset{2}{\prod}(x,y) = \{\pi \in \overset{1}{\prod} \mid \ & \exists M \ \in \mathcal{M}(x,y) \ \textit{such that}\ \forall\, w \in \ M, \\ & w^T \cdot \nabla^2_{yy} l(x,y;\pi) \cdot w \geq 0\} \end{aligned} \tag{14}$$

The *singular* multiplier sets $\prod^1_0(x,y)$ and $\prod^2_0(x,y)$ can be defined in the same way, except that the Lagrangean $l(x,y;\pi)$ is replaced by the *singular* Lagrangean

$$l_0(x,y;\pi) = l(x,y;\pi) - f(x,y) = \pi^T \cdot g(x,y).$$

Another multiplier set of interest is the *augmentable* multiplier set $\prod^a(x,y)$ related to the augmented Lagrangean

$$L(x,y;\pi^a,r) = l(x,y;\pi^a) + \frac{1}{2r}\sum_{i=1}^{m}([\pi_i^a + r g_i(x,y)]_+^2 - [\pi_i^a + r g_i(x,y)]^2)$$

where

$$[n]_+ = \max\{n, 0\}.$$

The *augmentable* multiplier set $\prod^a(x,y)$ consists of all the multiplier vectors such that for sufficiently large r, (y, π^a) is a local saddle point of $L(x,\cdot,\cdot,r)$ at (y,π). Rockafellar (1984) studies the properties of the multiplier sets and reveals the role they play in the directional differential property of the marginal function $v(x)$.

Proposition 5 *(Rockafellar 1974, 1984)* $\prod_0^1(x,y) = \{0\}$ *is equivalent to the Mangasarian-Fromovitz constraint qualification holding for* (P_x) *at* y. *A sufficient condition that* y *is a local optimal solution to* (P_x) *is that* $\prod^a(x,y) \neq \emptyset$. □

Proposition 6 *(Theorem 2, Rockafellar 1984) The second-order multiplier sets* $\prod^2(x,y)$ *and* $\prod_0^2(x,y)$ *are closed. If* $\prod_0^2(x,y) = \{0\}$, *then* $\prod^2(x,y)$ *is nonempty and compact for a locally optimal solution* y *to* (P_x). *In particular, a necessary condition for the lacal optimality of* y *in* $(P_x o$ *is the existence of a vector* $\pi \in \prod^2(x,y)$ *if* $\prod_0^2(x,y) = \{0\}$ *is satisfied.* □

To ensure that $O(x)$ is a nonempty compact set for each x such that $v(x) < \infty$ and $v(\cdot)$ becomes nowhere $-\infty$, the following *inf-boundedness* assumption is made to simplify the analysis:

> For each $x' \in X$ and $\alpha \in \mathbf{R}$, there exists an $\epsilon > 0$ such that the set
>
> $$\{(x,y) \in X \times Y \mid \mid x - x' \mid \leq \epsilon,\ f(x,y) \leq \alpha,\ g_i(x,y) \leq \epsilon,\ i = 1, \cdots, m\}$$
>
> is bounded.

The following result gives the upper estimate for the *Clarke directional derivative* and lower estimate for the *Hadamard directional derivative* of $v(x)$.

Proposition 7 *(Theorem 5, Rockafellar 1984) Let* $Y(x) \neq \emptyset$ *and every* $y \in O(x)$ *satisfy the constraint qualification* $\prod_0^2(x,y) = \{0\}$. *Then* $v(\cdot)$

is finite and Lipschitz continuous on a neighborhood of x and for every $k \in \mathbf{R}^{|X|}$

$$v^0(x;k) \leq \max_{y \in O(x)} \max_{\pi \in \prod^2(x,y)} k^T \cdot \nabla_x l(x,y;\pi), \tag{15}$$

If, for $\forall y \in O(x)$, $\prod^a(x,y) \neq \emptyset$, then there exists $y' \in O(x)$ such that

$$v_+(x;k) \geq k^T \cdot \nabla_x l(x,y;\pi) \quad \text{for } \forall \pi \in \overset{a}{\prod}(x,y') \tag{16}$$

Consequently

$$v_+(x;k) \geq \inf_{y \in O(x)} \sup_{\pi \in \prod^a(x,y)} k^T \cdot \nabla_x l(x,y;\pi) \tag{17}$$

where the lower and upper Hadamard directional derivative $v_+(x;k)$ and $v^+(x;k)$ are defined respectively by

$$v_+(x;k) = \liminf_{\substack{k' \to k \\ t \downarrow 0}} \frac{v(x+tk') - v(x)}{t},$$

$$v^+(x;k) = \limsup_{\substack{k' \to k \\ t \downarrow 0}} \frac{v(x+tk') - v(x)}{t}.$$

□

Next we present the main result of this section.

Theorem 1 *Let $x_0 \in X$ and $Y(x_0) \neq \emptyset$. If every $y \in O(x_0)$ satisfies the constraint qualification $\prod_0^2(x_0,y) = \{0\}$, then*

$$\partial v(x_0) \subseteq co.\{\nabla_x l(x_0,y;\pi) \mid y \in O(x_0),\ \pi \in \overset{2}{\prod}(x_0,y)\}. \tag{18}$$

In addition, if the marginal map $O(\cdot)$ is l.s.c. at x_0 and $\prod^a(x_0,y) \neq \emptyset$ for all $y \in O(x_0)$, then

$$\partial v(x_0) = co.\{\nabla_x l(x_0,y;\pi) \mid y \in O(x_0),\ \pi \in \overset{2}{\prod}(x_0,y)\}, \tag{19}$$

where co. stands for convex hull.

Proof. For any $\xi \in \partial v(x_0)$, we have

$$\begin{aligned} \xi^T \cdot k &\leq v^0(x_0; k) \quad (\textit{by definition}) \\ &\leq \max_{y \in O(x_0)} \max_{\pi \in \prod^2(x_0, y)} k^T \cdot \nabla_x l(x_0, y; \pi) \quad (\textit{by } (15)) \end{aligned}$$

Thus

$$\xi^T \cdot k \leq \max\{\eta^T \cdot k \mid \eta \in \bigcup_{y \in O(x_0)} \bigcup_{\pi \in \prod^2(x_0, y)} \nabla_x l(x_0, y; \pi)\}.$$

By Theorem 32.2 in (Rockafellar, 1970), we have

$$\xi^T \cdot k \leq \max\{\eta^T \cdot k \mid \eta \in co.\{ \bigcup_{y \in O(x_0)} \bigcup_{\pi \in \prod^2(x_0, y)} \nabla_x l(x_0, y; \pi)\}\}.$$

This implies that

$$\xi \in co.\{\nabla_x l(x_0, y; \pi) \mid y \in O(x_0),\ \pi \in \prod^2(x_0, y)\} \tag{20}$$

by Theorem 13.1 (Rockafellar 1970). Therefore the first part of the assertion is proved.

If the marginal map $O(\cdot)$ is l.s.c. at x_0 and $\prod^a(x_0, y) \neq \emptyset$, then for any $\bar{y} \in O(x_0)$, there exists a sequence $k^j \to k$ and $t_j \downarrow 0$ such that

$$\begin{aligned} &x_0 + t_j k^j \to x_0,\ v(x_o + t_j k^j) \to v(x_0) \ \textit{and} \\ &\qquad \frac{v(x_0 + t_j k^j) - v(x_0)}{t_j} \to v_+(x_0; k) \end{aligned}$$

without loss of generality. Then, by Theorem 7 (Rockafellar 1984),

$$\begin{aligned} v^+(x_0; k) &\geq v_+(x_0; k) \\ &\geq \sup_{\pi \in \prod^a(x_0, \bar{y})} k^T \cdot \nabla_x l(x_0, \bar{y}; \pi). \end{aligned}$$

Hence

$$v^+(x_0; k) \geq \sup_{y \in O(x_0)} \sup_{\pi \in \prod^a(x_0, y)} k^T \cdot \nabla_x l(x_0, y; \pi). \tag{21}$$

Therefore, for any

$$\xi \in co.\{k^T \cdot \nabla_x l(x_0, y; \pi) \mid y \in O(x_0),\ \pi \in \overset{a}{\prod}(x_0, y)\},$$

we have

$$\begin{aligned} \xi^T \cdot k &\leq v^+(x_0; k) \quad (by\ (21)) \\ &\leq v^0(x_0; k) \quad for\ all\ k \in X^*. \end{aligned}$$

This leads to

$$\xi \in \partial v(x_0). \tag{22}$$

Combining with (20), we obtain

$$\partial v(x_0) = co.\{\nabla_x l(x_0, y; \pi) \mid y \in O(x_0),\ \pi \in \overset{a}{\prod}(x_0, y)\}. \tag{23}$$

□

If further assumptions are made about the marginal map and/or the second-order multiplier set, tighter results can be obtained.

Corollary 1 *If $O(x_0) = \{y_0\}$ is a singleton, that is, the second level problem provides a unique optimal solution for x_0 and $\prod^a(x_0, y_0) \neq \emptyset$, then*

$$\partial v(x_0) = co.\{\nabla_x l(x_0, y_0; , \pi) \mid \pi \in \overset{a}{\prod}(x_0, y_0)\} \tag{24}$$

□

It is noted that the condition $\prod^a(x_0, y_0) \neq \emptyset$ can be relaxed to the condition $\prod^2(x_0, y_0) \neq \emptyset$ (Gauvin and Dubeau 1982).

Corollary 2 *If the second level problem has unique optimal solution y_0 and unique optimal multiplier π_0 at x_0, that is, $O(x_0)$ and $\prod^a(x_0, y_0)$ $(= \prod^2(x_0, y_0))$ are singletons, then the marginal function $v(x)$ is differentiable in the sense of Gâteaux differentiability, and the gradient of $v(\cdot)$ at x_0 is*

$$\begin{aligned} \nabla v(x_0) &= \nabla_x l(x_0, y_0); \pi^0) \\ &= \nabla_x f(x_0, y_0) + \pi^{0T} \cdot \nabla_x g(x_0, y_0) \end{aligned} \tag{25}$$

□

With these differential properties of the marginal function available, we can proceed to derive the optimality conditions of the (BLPP).

4 First-order Optimality Conditions for the General BLPP

The study of optimality conditions is an important issue in optimization since it is crucial to the development of efficient algorithms for the problems under study. Danskin (1966), Bram (1966) and Schmitendorf (1977) study the max-min problem, a special case of the BLPP, and obtain necessary and sufficient optimality conditions for the problem. For the general BLPP, Bard (1984) proposes a set of necessary conditions by considering an equivalent infinitely constrained optimization problem defined as follows:

$$\begin{aligned} \min\ & F(x,y) \\ s.t.\ & f(x,y) - f(x,z) \le 0 \quad \textit{for all } z \in \{z \mid g(x,z) \le 0\} \\ & g(x,y) \le 0. \end{aligned}$$

However, as Clark and Westerberg (1988) point out by an counterexample, those conditions are not valid. The establishment of the optimality conditions for the BLPP remained an open problem for a long time.

Fortunately, by taking advantage of the particular formulation of the BLPP and the (generalized) differentiability of the marginal function, we are in the position to establish the first-order necessary optimality condition for the BLPP in the framework of nonsmooth optimization. The sought result follows.

Theorem 2 *If (x_0, y_0) solves the BLPP $(P1)$, $\prod^2(x_0, y_0) \neq \emptyset$, $\prod_0^2(x_0, y_0) = \{0\}$ and one constraint qualification (for example, Cottle constraint qualification) holds for $(P2)$ at (x_0, y_0), then there exist a scalar $\alpha \in \mathbf{R}$ and a vector $\theta \in \mathbf{R}^m$ such that*

$$\nabla_x F(x_0,y_0) + \alpha \cdot [\nabla_x f(x_0,y_0) - \partial v(x_0)] + \theta^T \cdot \nabla_x g(x_0,y_0) \ni 0 \quad (26)$$

$$\nabla_y F(x_0,y_0) + \alpha \cdot \nabla_y f(x_0,y_0) + \theta^T \cdot \nabla_y g(x_0,y_0) = 0, \quad (27)$$

$$\theta_i \cdot g_i(x_0,y_0) = 0, \quad i = 1, \cdots, m \quad (28)$$

$$f(x_0,y_0) - v(x_0) = 0, \quad (29)$$

$$g(x_0,y_0) \le 0, \quad (30)$$

$$\theta \ge 0. \quad (31)$$

where $\partial v(x_0)$ may be computed according to one of the formulae (23) *and* (24).

Proof. It follows from a straightforward application of the Karush-Kuhn-Tucker conditions to the problem $(P2)$. □

The following corollary sharpens the above result when certain conditions are satisfied.

Corollary 3 *Let (x_0, y_0) solve the BLPP $(P2)$ and a constraint qualification for $(P2)$ hold at (x_0, y_0), and $\prod_0^2(x_0, y_0) = \{0\}$ as well. If the second level problem has a unique optimal solution and unique optimal multiplier π_0, for $x = x_0$, then there exist $\alpha \in \mathbf{R}$ and $\theta \in \mathbf{R}^m$ such that*

$$\nabla_x F(x_0, y_0) + (\theta - \alpha\pi_0)^T \cdot \nabla_x g(x_0, y_0) = 0, \tag{32}$$

$$\nabla_y F(x_0, y_0) + \alpha \cdot \nabla_y f(x_0, y_0) + \theta^T \cdot \nabla_y g(x_0, y_0) = 0, \tag{33}$$

$$\theta_i \cdot g_i(x_0, y_0) = 0, \quad i = 1, \cdots, m \tag{34}$$

$$f(x_0, y_0) - v(x_0) = 0, \tag{35}$$

$$g(x_0, y_0) \leq 0, \tag{36}$$

$$\theta \geq 0. \tag{37}$$

Proof. By replacing for $\partial v(x_0)$ in (26) by formulae (25) we get the result.

□

In order to have an insight into the optimality conditions stated above, let us apply it to the counterexample for Bard's condition provided by Clark and Westerberg (1988).

$$\begin{aligned} &\min_x x - 4y \\ &\quad \min_y y \\ &\qquad s.t. \quad g_1 = -x + y \leq 0, \\ &\qquad\quad g_2 = 2x + 5y - 108 \leq 0, \\ &\qquad\quad g_3 = 2x - 3y + 4 \leq 0. \end{aligned}$$

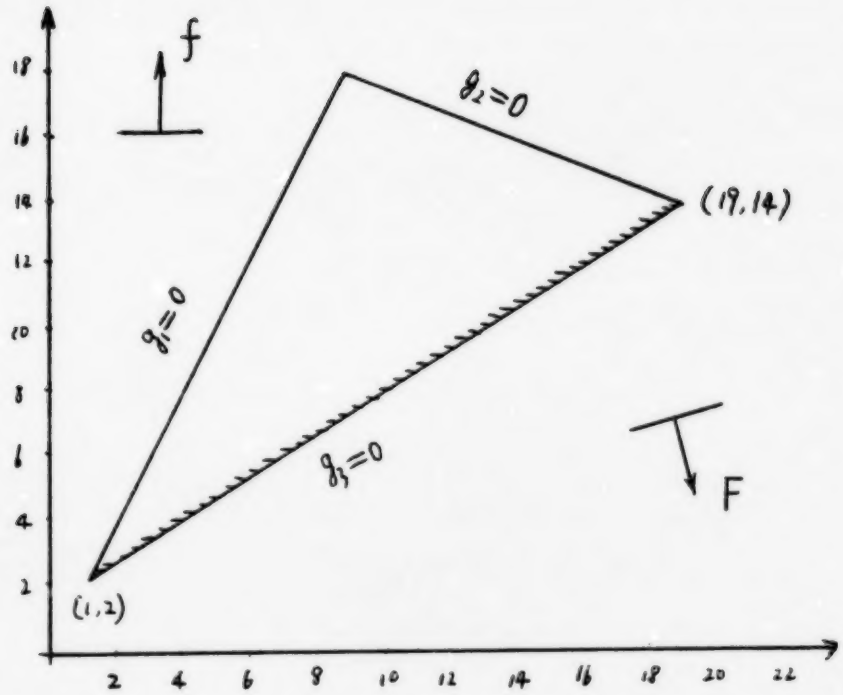

Figure 1: Problem geometry of the example

The problem geometry is shown in Figure 1. The global minimum occurs at (19,14) and the global maximum at (1,2). At the global minimum point (19,14) where Bard's condition fails, the optimality condition reads:

$$\begin{aligned} 1 - 2\theta_1 + 2\theta_2 + 2\theta_3 + 2\alpha\pi_1 - 2\alpha\pi_2 - 2\alpha\pi_3 = 0, \\ -4 + \alpha + \theta_1 + 5\theta_2 - 3\theta_3 = 0, \\ \theta_1 = 0, \quad \pi_1 = 0, \\ \theta \geq 0, \quad \pi \geq 0. \end{aligned}$$

The multipliers satisfying the system are

$$\begin{aligned} \alpha &= 1, \quad , \\ \pi &= (0, 2, 0), \\ \theta &= (0, \frac{15}{16}, \frac{9}{16}). \end{aligned}$$

Figure 2 shows the cone produced by the gradients of the active constraints and the vector $-(0, \nabla_y f)$ where ∇F lies at point (19,14). This is exactly the requirement of the optimality conditions.

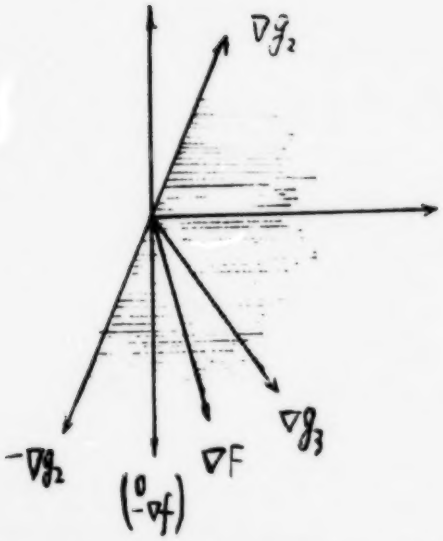

Figure 2: The necessary conditions at (19,14)

The "theorem of alternative " result is given to obtain further insight to the problem structure of the BLPP.

Theorem 3 *If (x_0, y_0) solves the BLPP $(P1)$, then there is no solution to the following system of inequalities:*

$$(w, z)^T \cdot \nabla F(x_0, y_0) < 0, \tag{38}$$

$$z^T \cdot \nabla_y f(x_0, y_0) \leq 0, \tag{39}$$

$$w^T \cdot \nabla_x g_i(x_0, y_0) = 0, \quad i =\in I(x_0, y_0) \tag{40}$$

$$z^T \cdot \nabla_y g_i((x_0, y_0) \leq 0, \quad i \in I(x_0, y_0) \tag{41}$$

where $(w, z) \in \mathbf{R}^{|X|+|Y|}$.

Proof. It is noted that (32) may be rewritten as

$$\nabla_x F(x_0, y_0) + \theta^T \cdot \nabla_x g(x_0, y_0) + \alpha\beta^T \cdot (-\nabla_x g(x_0, y_0)) = 0. \tag{42}$$

Applying *Farkas'* lemma to (42) implies that the set

$$\begin{aligned} & \{w \mid w^T \cdot \nabla_x F(x_0, y_0) < 0,\ w^T \cdot \nabla_x g(x_0, y_0) \leq 0,\ w^T \cdot \nabla_x(-g(x_0, y_0)) \leq 0\} \\ = \ & \{w \mid w^T \cdot \nabla_x F(x_0, y_0) < 0,\ w^T \cdot \nabla_x g(x_0, y_0) = 0\} \end{aligned}$$

has no solution. The result is proved by using *Farkas'* lemma again to (33). □

This assertion bears some analogy to those in standard single level programs except that the inequality constraints $g(x,y) \leq 0$ is treated as equality ones by the level one player. Therefore the set determined by (39) -(41) defines the cone of feasible direction of the BLPP, while the set by (38) – (41) is the descent feasible direction cone. It may also be interpreted as follows: if the solution is achieved, there is no descent feasible direction for the level one player, in the sense that all constraints behave like equality constraints; and there is no feasible direction along which the second level objective would not increase and the first level objective decrease at the same time. The result above is significant not only for solving the long open problem of finding the true cone of feasible direction of the BLPP, but also for the development of efficient algorithms.

It is interesting to note that one can relate the BLPP to another single level problem based on the necessary optimality conditions stated above.

Corollary 4 *If (x_0, y_0) solves the BLPP (P1), then there exist $\alpha \in \mathbf{R}$, $\pi \in \mathbf{R}^m$ and $\theta \in \mathbf{R}^m$ such that $(y_0, \alpha, \pi, \theta)$ solve the following problem when $x = x_0$ fixed:*

$$
\begin{aligned}
(P3) \qquad & \min f(x,y) && (43)\\
& s.t. \qquad \nabla_x F(x,y) + (\theta - \alpha\pi)^T \cdot \nabla_x g(x,y) = 0 && (44)\\
& \nabla_y F(x,y) + \alpha \cdot \nabla_y f(x,y) + \theta^T \cdot \nabla_y g(x,y) = 0, && (45)\\
& \theta_i \cdot g_i(x,y) = 0, \quad i = 1, \cdots, m && (46)\\
& g(x,y) \leq 0, && (47)\\
& \theta \geq 0, \ \pi \geq 0. && (48)
\end{aligned}
$$

The solution to (P3) provides a lower bound to (P2). □

5 Applications to Some Special BLPPs

In this section, the emphasis is focused on the convex BLPP, which receives much attention in the current research, and the min-max problem, which, being of importance by its own, can be considered as a special case of the BLPP. From the unified framework of the BLPP

discussed earlier, more simpler but more powerful conditions for the convex BLPP and more general results for the min max problem can be obtained.

5.1 The Convex BLPP

In the formulation of the BLPP, the second level problem (P_x) will define a convex programming problem, provided $f(x,\cdot)$ and $g(x,\cdot)$ are convex functions on y. The BLPP with convex second level problem is called the convex BLPP. The convex BLPP is one of the most important classes in the BLPP family both in the theoretical and practical sense. It is quite remarkable that for the convex BLPP the optimality conditions are much simpler and more powerful comparing with those for the general BLPP.

Theorem 4 *Let $f(x,\cdot)$ and $g(x,\cdot)$ be convex functions on y and the conditions of the Corollary 3 be satisfied. If, in addition, a constraint qualification holds for the second level problem at x_0, then there exist $\alpha \in \mathbf{R}$, $\theta \in \mathbf{R}^m$ and $\pi \in \mathbf{R}^m$ such that*

$$\nabla F(x_0,y_0) + (\theta - \alpha\pi)^T \cdot \nabla g(x_0,y_0) = 0, \tag{49}$$

$$\theta_i \cdot g(x_0,y_0) = 0,\ \pi_i \cdot g(x_0,y_0) = 0, \quad i = 1,\cdots,m. \tag{50}$$

$$g(x_0,y_0) \le 0, \tag{51}$$

$$\theta \ge 0,\ \pi \ge 0. \tag{52}$$

where the gradient are taken with respect to all variables.

Proof. The conditions (35) and (36) in Corollary 3 imply that $y_0)$ is the optimal solution to (P_{x_0}). By the convexity of $f(x,\cdot)$ and $g(x,\cdot)$, (35) and (36) are equivalent to the KKT conditions for (P_{x_0}). That is, there exists $\gamma \in \mathbf{R}^m$, $\gamma \ge 0$, such that

$$\nabla_y f(x,y) + \gamma^T \nabla_y \cdot g(x,y) = 0, \tag{53}$$

$$\gamma_i \cdot g_i(x,y) = 0, \quad i = 1,\cdots,m. \tag{54}$$

$$g(x,y) \le 0. \tag{55}$$

From (53), we get

$$\nabla_y f(x,y) = -\gamma^T \cdot \nabla_y g(x,y). \tag{56}$$

By replacing for $\nabla_y f(x,y)$ in (33) by (56), there is,

$$\nabla_y F(x,y) + (\theta - \alpha\gamma)^T \cdot \nabla_y g(x,y) = 0. \tag{57}$$

According to the uniqueness of the multipliers of (P_x), γ must be identical to π. Combining (32) and (57) we obtain (49). This completes the proof.

□

Similarly, the cone of the feasible direction for the convex BLPP can be characterized by the following corollary.

Corollary 5 *If (x,y) solves the convex BLPP, and the conditions in Corollary 3 are satisfied, then there is no solution to the following system:*

$$\begin{aligned} w^T \cdot \nabla F(x,y) &< 0, \\ w^T \cdot \nabla g_i(x,y) &= 0, \quad i \in I(x,y). \end{aligned}$$

where w are $\mathbf{R}^{|X|+|Y|}$ dimension vectors.

The information from this corollary tells one that, for the convex BLPP, all the active constraints behave like equality constraints even though they may be inequality constraints. This is important for constructing algorithms for convex BLPP.

5.2 Min-Max Problem

The well studied min-max problem is defined as follows

$$(P_4) \qquad \min_{x \in X} \max_{y \in Y} \phi(x,y)$$

where $X = \{x \mid g(x) \leq 0\}$ and Y is a fixed topological space.

Obviously, the min-max problem is a special case of the BLPP, since if $F(x,y) = -f(x,y)$, the problem (P_1) leads to the following generalized min-max problem:

$$(P_5) \qquad \min_{x \in S_x} \max_{y in Y(x)} F(x,y)$$

where $S_x = \{x \in X \mid (x,y) \in S\}$ is the projection of $S = \{(x,y) \in X \times Y \mid l(x,y) \leq 0,\ g(x,y) \leq 0\}$ onto X.

Danskin (1966), Bram (1966) and Schmitendorf (1977) study the optimality conditions of the min-max problem of form (P_4). From bilevel viewpoint, their results can be generalized to the more general problem (P_5) in the framework stated above.

Corollary 6 *Let (x,y) solve the problem (P_5). Then there exist $\alpha \in \mathbf{R}$, $\theta \in \mathbf{R}^m$ and $\eta \in \mathbf{R}^m$ such that*

$$(1+\alpha) \cdot \nabla_x F(x,y) + \eta^T \cdot \nabla_x l(x,y) + \theta^T \cdot \nabla_x g(x,y) - \alpha \cdot \partial v(x) \ni 0, \quad (58)$$

$$(1+\alpha) \cdot \nabla_y F(x,y) + \eta^T \cdot \nabla_y l(x,y) + \theta^T \cdot \nabla_y g(x,y) = 0, \quad (59)$$

as well as the complementarity and feasibility conditions.

□

If $l(x,y) = l(x)$ and $Y(x) = Y$, the problem (P_5) becomes (P_4). Then (58) and (59) reduce to

$$\eta^T \cdot \nabla_x l(x,y) - \alpha \cdot \partial v(x) \ni 0, \quad (60)$$

$$(1+\alpha) \cdot \nabla_y F(x,y) = 0. \quad (61)$$

which is equivalent to the result given by Schmitendorf (Theorem 1, Schmitendorf (1977)).

6 Conclusion

The study of the optimality condition is an important issue in optimization both in theoretical and practical aspects. In this article we propose a general formulation of the bilevel problem and provide a positive answer to the problem of finding the optimality condition and the cone of the feasible direction for the general BLPP, which has been unanswered until now. The results are significant not only for enabling one to gain insights to the problem but also for providing the basis for developing efficient algorithms for the BLPP.

One approach motivated by these ideas is to design *feasible direction methods* for a class of BLPP problems, which take full advantage of the theory developed here. Such algorithms are the subject of a separate paper. Other methods which readily come to mind for the BLPP are applications of the augmented Lagrange approach or penalty function approaches, since they provide a natural way to deal with the constraint involved the marginal function. By virtue of the properties of the problem formulation and the optimality condition developed in the paper, it is definitely possible to also adapt approximation techniques, especially for the approximation of the marginal function. In the authors' opinion, they may be one of the most promising approaches to solve the BLPP problem. Nevertheless, the finding of the global optimal solution for the general BLPP remains a tough problem. It is evident from the formulation that the global solution of the problem depends on properties of the marginal function, such as the generalized convexity and so on. Hopefully, for some special BLPP problems, such as the BLPP with linear second level problems, one can derive satisfactory solution methods for finding the global optimal solution.

References

[1] Aiyoshi, E. and Shimizu K. (1981), Hierarchical Decentralized System and Its New Solution by a Barrier Method, *IEEE Transactions on Systems, Man and Cybernetics* SMC-11, 444-449.

[2] Aubin and Ekeland (1984), Applied Nonlinear Analysis, *John Wiley and Sons* .

[3] Bard, J.F. (1983), An Algorithm for Solving the General Bilevel Programming Problem, *Mathematics of Operations Research* 8, 260-272.

[4] Bard, J.F. (1984), Optimality Conditions for the Bilevel Programming Problem, *Naval Research Logistics Quarterly* 31, 13-26

[5] Bard, J.F. and J.E. Falk (1982), An Explicit Solution to the Multilevel Programming Problem, *Computers and Operations Research* 9, 77-100.

[6] Bard, J.F. and J. T. Moore (1990), A Branch and Bound Algorithm for Bilevel Programming Problem, *SIAM on Scientific and Statistical Computing* 11, 281-292.

[7] Ben-Ayed, O., D.E. Boyce and C.E. Blair,III (1988), A General Bilevel Linear Programming Formulation of the Network Design Problem, *Transportation Research B* 22, 311-318.

[8] Benson, H.P. (1989), On the Structure and Properties of a Linear Multilevel Programming Problem, *Journal of Optimization Theory and Applications* 60, 353-373.

[9] Bialas, W.F. and M.H. Karwan (1982), On Two-level Optimization, *IEEE Transactions on Automatic Control* AC-27, 211-214.

[10] Bialas, W.F. and M.H. Karwan (1984), Two-level Linear Programming, *Management Science* 30, 1004-1020.

[11] Bram, J. (1966), The Lagrange Mulliplier Theorem for Max-Min Problem with several constraints, *SIAM J. Applied Mathematics* 14, 665-667.

[12] Candler, W. and R. Townsley (1982), A Linear Two-level Programming Problem, *Computers and Operations Research* 9, 59-76.

[13] Chen, C.I. and J.B. Cruz Jr. (1972), Stackelberg Solution for Two-Person Games with Biased Information Patterns, *IEEE Transactions on Automatic Control* AC-17, 791-798.

[14] Clark, P.A. and A.W. Westerberg (1988), A Note on the Optimality Conditions for the Bilevel Programming Problem, *Naval Research Logistics* 35, 414-418.

[15] Clarke, F.H. (1983), *Nonsmooth Analysis and Optimization* Wiley-Inter science, New York.

[16] Danskin, J.M. (1966), The Theory of Max-Min, with Applications, *J. SIAM Applied Mathematics* 14, 641-664.

[17] Falk, J.E. (1973), A Linear Max-Min Problem, *Mathematical Programming* 5, 169-188.

[18] Fiaco, A.V. (1976), Sensitivity Analysis for Nonlinear Programming Using Penalty Methods, *Mathematical Programming* 10, 287-311.

[19] Florian, M and Y. Chen (1991), A Bilevel Programming Approach to Estimating O-D Matrix by Traffic counts, *Publication CRT-750, Université de Montréal, Montréal, Canada.*

[20] Fortuny-Amat, J and B. McCarl (1981), A Representation of a Two-Level Programming Problem, *Journal of the Operational Research Society* 32, 783-792.

[21] Friesz, T.L., R.L. Tobin, H. Cho and N.J. Mehta (1990), Sensitivity Analysis Based Heuristic Algorithms for Mathematics Programs with Variational Inequality constraints, *Mathematical Programming* 48, 265-284.

[22] Gauvin, J. and F. Dubeau (1982), Differential Properties of the Marginal Function in Mathematical Programming, *Mathematical Programming Study* 19, 101-119.

[23] Gauvin, J. and G. Savard (1990), The Steepest Descent Method for the Nonlinear Bilevel Programming Problem, *Technical Report* .

[24] Hansen, P., B. Jaumard and G. Savard (1990), New Branching and Bounding Rules for Linear Bilevel Programming, *GERAD Publication 0711-2440* .

[25] Hogan, W.W. (1973) Point-to-Set Maps in Mathematical Programming, *SIAM Review* 15, 591-603.

[26] Jeroslow, R.G. (1985), The Polynomial Hierarchy and a Simple Model for Competitive Analysis, *Mathematical Programming* 32, 146-164.

[27] Judice, J.J. and A.M. Faustino (1988), The Solution of the Linear Bilevel Programming Problem by Using the Linear Complementary Problem, *Investigacão Operacional* 8, 77-95.

[28] Judice, J.J. and A.M. Faustino (1989), A Sequential LCP Method for Bilevel Linear Programming, *Research Report* .

[29] Kolstad, C.D. and L.S. Lasdon (1990), Derivative Evaluation and Computational Experience with Large Bilevel Mathematical Programming, *Journal of Optimization and Applications* 65, 485-499.

[30] LeBlanc, L. and D.E. Boyce (1986), A Bilevel Programming Algorithm for Exact Solution of the Network Design Problem with User-Optimal Flows, *Transportation Research B* 20, 259-265.

[31] Loridan, P. and J. Morgan (1989a), A Theoretical Approximation Scheme for Stackelberg Problems, *Journal of Optimization and Applications* 61, 95-111.

[32] Loridan, P. and J. Morgan (1989b), New Results on Approximate Solutions in Two -Level Optimization, *optimization* 20, 819-836.

[33] Morcotte, P. (1986), Network Design Problem with Congestion Effects: A Case of Bilevel Programming, *Mathematical Programming* 34, 142-162.

[34] Rockafellar, R.T. (1970), Convex Analysis. *Princeton University Press* .

[35] Rockafellar, R.T. (1974), Augmented Larange Multiplier Functions and Duality in Nonconvex Programming, *SIAM Journal on Control* 12, 268-285.

[36] Rockafellar, R.T. (1984), Directional Differentiability of the Optimal Value Function in a Nonlinear Programming Problem, *Mathematical Programming Study* 21, 213-226.

[37] Schmitendorf, W.E. (1977), Necessary Conditions and Sufficient Conditions for Static Max-Min Problems, *Journal of Mathematical Analysis and Applications* 57, 683-693.

[38] Spiess, H. (1990), A Gradient Approach for the O-D Matrix Adjustment Problem, *CRT Publication 693, Université de Montreal, Montreal, Canada.*

[39] Stackelberg Von, H. (1952), *The Theory of the Market Theory* , Oxford University Press, Oxford, England.